

Technical Report:

Analysis of the Psychometric Structure of the LSC Surveys

David B. Flora & A. T. Panter

L. L. Thurstone Psychometric Lab

University of North Carolina at Chapel Hill

The analyses of the psychometric structure of the Local Systemic Change (LSC) surveys followed a conservative course. Although latent variables had been defined *a priori*, an item-level exploratory factor analysis was desirable for the purpose of identifying which, if any, individual items could be eliminated from the lengthy LSC survey instruments. Therefore, the analysis proceeded according to a series of stages—first, an exploratory factor analysis was performed, and then two rounds of confirmatory factor analysis followed. The entire K-8 science data set ($N = 6156$) was randomly divided into three smaller data sets, each consisting of approximately 2050 observations. These three smaller data sets were created so that each stage of the analyses could be performed on an independent sample. The observations in the K-8 mathematics data set were also subdivided into three random samples, while the smaller 7-12 mathematics data set was divided into only two random samples, one for exploratory factor analysis and the other for confirmation.

Stage 1: Exploratory factor analysis and item analysis

The first stage of the analyses consisted of exploratory factor analyses and reliability analyses. Horizon Research, Inc defined five domains of items, or “Core Evaluation Questions,” *a priori*. With the exception of one domain, separate exploratory

factor analyses and reliability analyses were performed on each domain.¹ The purpose of the exploratory factor analyses was to identify the items within a domain that seemed to be measuring the same latent variable, or construct. Once a group of items was determined to measure a particular construct, reliability statistics (item-total correlations and Cronbach's coefficient alpha) were calculated for each set of items, or scale, comprising a construct. These reliability statistics provide an indication of how well a set of items measures a single construct. Additionally, these reliability estimates can be shown to be mathematically equivalent to an indication of how well measurement models implied by the factor analyses will remain stable across new samples.

It is important to note that principal axis factor analysis was utilized rather than principal components analysis in these analyses. Although principal axis factor analysis and principal components analysis are nearly identical procedures mathematically, their purposes are fundamentally distinct. Principal components analysis takes a set of variables and defines a new set of variables (completely recapturing all of the variability among the original set of variables). Thus, this approach assumes that there is zero measurement error associated with the original variables. Principal axis factoring, on the other hand, allows for the possibility that the manifest variables may have error associated with their ability to define the latent variables, or constructs, which are created with the analysis. In the principal axis model, each manifest variable is composed of two elements: a communality, the portion of the variable that measures the latent construct, and a uniqueness, the portion of the variable that either measures a different construct or is error (or some combination of error and another construct). The communality

¹ One domain consisted of demographic items and was omitted from all of the analyses described herein.

estimates used for the present principal axis procedures consisted of the squared multiple correlation of a given variable with all other variables in the set of manifest variables.²

Once a decision was reached about the proper number of factors to extract for a given content domain, these factors were rotated according to an oblique, rather than orthogonal, procedure. Factor rotation is performed so that each factor, or latent construct, is clearly defined by one set of original variables, which have high factor loadings (i.e., high correlations with the construct), while all other manifest variables have low factor loadings. Oblique rotation allows the latent constructs to be correlated with each other, whereas orthogonal rotation leads to latent constructs that have zero correlation between them. The Promax rotation was the oblique rotational procedure used with the present analyses.³

When constructing a survey, researchers often have an item pool that is conceptually related at one level of abstraction, but at another level of thought the pool could possibly be divided into two or more sub-pools of items. Items within these smaller pools will share substantive meaning but the substantive interpretations could vary across the sub-pools. Exploratory factor analyses can help researchers define how item subsets from a survey may come together in scales that define more specific constructs. Furthermore, exploratory factor analyses can help identify which items in a survey do not seem to be measuring the construct they were originally intended to

² Squared multiple correlations were chosen for prior communality estimates simply because they are viewed as an estimate with excellent properties for this analysis. Examples of other prior communality estimates include average correlation and highest correlation. The choice of prior communality estimate creates little cause for concern because factor analytic results tend to be virtually identical regardless of what type of communality estimate is used, particularly for item sets with several variables.

³ Promax rotation was chosen due to its relative mathematical simplicity. As is the case with choice of communality estimate, results tend to be stable across different oblique rotations.

measure. For these reasons, the item behavior on the LSC k-8 science survey was examined using exploratory factor analysis.

Because the exploratory factor analytic results may capitalize on chance aspects of the item associations (see Nunnally and Bernstein, 1994), an important step was to conduct reliability analyses once potential scales were identified. The reliability statistics provide another form of evidence of how well the items within a scale define a construct. Although factor loadings for individual items indicate how strongly a given item is associated with a construct, factor loadings can be unstable. Once a scale is established from a set of items, the correlations between individual items within the set and the total, or summed, score of all items defining the scale provide stable estimates of how strongly individual items are associated with the scale's construct. Cronbach's coefficient alpha provides an estimate of the internal consistency of the entire scale. This estimate of internal consistency is mathematically equivalent to an estimate of a hypothetical correlation that would result if a parallel form of the scale were created and the correlation between the alternate form and the original form was calculated. The item-total correlations together with coefficient alpha provide robust estimates of the stability of the scale defining a construct.

Stage 2: Confirmation of structure

Once potential scales were identified to comprise constructs, a measurement model was implied for each domain. These measurement models were tested on the second random subsample of the data using a confirmatory factor analysis. In confirmatory factor analysis, latent variables are explicitly specified *a priori* according to which manifest variables should load onto them. These predetermined relations among

all of the variables (both manifest and latent) define a measurement model. The matrix of covariances between all manifest variables is analyzed using maximum likelihood procedures to test the extent to which the data accurately fit the specified model.

Various statistics provide information about how well the data fit the model. The most prominent of these statistics is a χ^2 statistic with degrees of freedom as a function of the number of parameters specified in the model. A statistically significant χ^2 often indicates that the data do not fit the model. However, large sample sizes, such as that employed here, inflate χ^2 statistics; thus, the χ^2 measure does not provide a good indication of how well the present data fit the models established from the exploratory factor analyses and reliability analyses. Rather, other measures of goodness of fit need to be consulted that do not rely so heavily on sample size. The present confirmatory analyses focused mainly on the root mean square error of approximation (Steiger & Lind, 1980) statistic (which is conceptually related to an average model residual) and the Comparative Fit Index (Bentler, 1990) statistic (which is conceptually related to the proportion of variability explained by the model) to determine whether these data fit the measurement models.

Upon defining these measurement models, it is sometimes the case that the measurement error for particular items (manifest variables) may be correlated. In the factor analytic model, the path from the latent variable to the manifest variables estimates each manifest variable's communality or relationship with the latent variable, while correlated measurement error terms between the manifest variables implies that these variables have correlated uniquenesses. These correlated uniquenesses could result from a participant response bias shared by items, or they could result from items measuring a

particular construct other than that defined in the present model. Hence, the present measurement models were tested in two ways; without correlated error terms and with particular error terms correlated. Error terms were allowed to correlate only when there was some conceptual rationale for inserting these paths into the models. For all domains, once measurement models with certain error terms allowed to correlate were specified, χ^2 difference tests revealed that these models fit the data statistically significantly better than models with no correlated errors.

Stage 3: Further confirmation

The results indicated that the constructs were quite similar in content across the three surveys. For constructs that were nearly identical, the items that were not shared by all three surveys were removed for the purpose of direct comparisons. Next, confirmatory factor analyses established that the psychometric structure defined in stages one and two existed in the remaining unanalyzed data sets.

References

- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238-246.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- Steiger, J. H., & Lind, J. M. (1980). *Statistically based tests for the number of common factors*. Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.